Reprinted from THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, Vol. 35, No. 8, 1201–1205, August, 1963

Detection, Discrimination, and the Loudness of Short Tones

C. DOUGLAS CREELMAN

Department of Psychology, The Johns Hopkins University, Baltimore 18, Maryland
(Received 26 April 1963)

Two experiments were conducted, with the same observers in each. A method of adjustment was first used to measure the relative loudness of short tonal signals at four different durations and four intensity levels. The second experiment measured the detectability of each duration which had been used in the first, and scales were constructed on the basis of the measurements. It was shown that the relative loudness of a short tone (the level to which it will be set to sound equally loud) can be predicted from the detectability. The relationship is an extremely simple one: the equal-loudness settings yield signals of equal detectability.

INTRODUCTION

OBSERVERS in this experiment performed two tasks. The first was loudness matching, where a short variable tone was adjusted in subjective loudness to equal a longer standard tone. This task was followed by a signal-detection experiment where the detectability of each signal duration which had been used in the matching task was measured. The results of each part of the experiment are in substantial agreement with previous studies. The important result was that the signal levels used by the subjects to match the loudness of tones of different durations could be predicted with considerable accuracy from the detection data.

SIGNAL DETECTION AND AMPLITUDE DISCRIMINATION

The four observers were run two at a time in a sound-treated room. The general procedure of the experiment has been described before.[1,2] A two-alternative temporal forced-choice procedure was used; after a warning light two flashes of a white light marked two temporal intervals. During one of the intervals a signal was presented monaurally to the observers in a background of continuous white noise at about 40 dB SL. During a subsequent answer period, marked by a third light, the observers indicated whether the signal had come in the first or the second interval. The choice of the interval for presentation of the signal was done automatically with equal probability for each alternative. A two-alternative trial took approximately 3.5 sec, and after each run of 100 trials the observers were given a brief rest before continuing. Knowledge of results was given by means of lights after each trial, and by the total, which was told them after each run.

The PDR-8 earphones were wired in parallel for monaural presentation, and were fed by a low-impedance amplifier output in an attempt to minimize transients.[3] The signals were segments of a 1000-cps sine wave generated by a General Radio 713-B beat-frequency audio oscillator, gated by an electronic switch so that they were turned on at a positive-going zero crossing of the waveform, and continued for an integral number of cycles.

Detection experiment. Psychophysical functions were constructed, plotting measured detectability as a function of signal voltage for signals of 5 durations: 5, 10, 40, 160, and 320 msec. Performance was measured in terms of percent correct responses, and converted to the detectability measure d' by means of tables.[4] Detectability thus measured is shown by the solid points in the functions for the four observers in Fig. 1. There are five sets of points for each observer, one for each signal duration. The longest, 0.32 sec, has its points plotted against the lowest signal voltages, as we would expect.

[1] D. M. Green, T. G. Birdsall, and W. P. Tanner, Jr., "Signal Detection as a Function of Signal Intensity and Duration," J. Acoust. Soc. Am. 29, 523–531 (1957).

[2] W. P. Tanner, Jr., "Application of the Theory of Signal Detectability to Amplitude Discrimination," J. Acoust. Soc. Am. 33, 1233–1244 (1961).

[3] R. C. Bilger, "Laboratory Facilities Employed in Psychophysical Memory Experiments," Tech. Memo. No. 72, Electronic Defense Group, Univ. of Michigan (1959).

[4] P. B. Elliott, "Tables of d'," Tech. Rept. No. 97, Electronic Defense Group, Univ. of Michigan (1959).

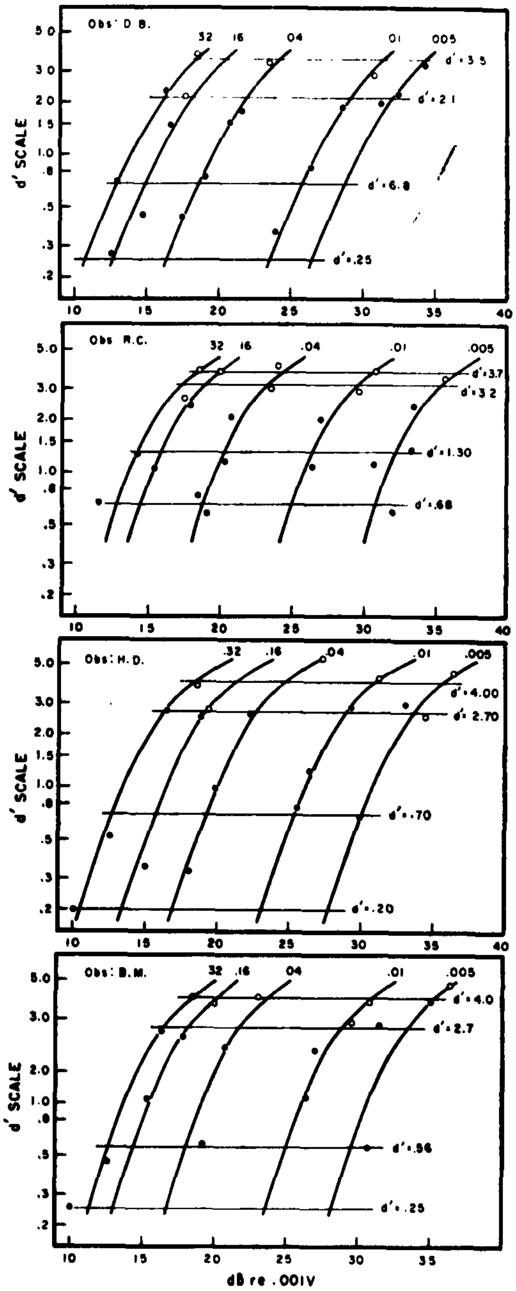

FIG. 1. Detectability scales for 5 signal durations for 4 observers. The abscissa is the signal voltage to the observers' earphones, in dB. The parameter at the top of the curves is the signal duration. The horizontal lines are the basis for prediction of loudness matching. Solid points indicate direct detectability measurements, and open circles indicate scale values obtained from amplitude discrimination.

The open circles show scale values estimated by amplitude discrimination.

Amplitude discrimination experiment. In addition to detection, larger amplitude signals, which could be detected nearly 100% of the time, were paired with lower amplitude signals in the detection range, also in a two-alternative format. On each trial the two signals were presented in random order. The task of the observer was to indicate for each trial whether the larger of the two signals had come first or second. Detectability measures were computed for amplitude discrimination from the obtained percent correct responses, in the same manner as in the detection experiment. This procedure, and some interesting results from it, has been reported by Pfafflin and Mathews.[5] The amplitude discrimination data were used to construct the rest of the psychophysical functions, shown in Fig. 1, by the additive procedure described below.

Discriminability scale for auditory signals. The scale of Fig. 1 was constructed in two steps. First the detectability of a signal presented in noise at a given voltage and duration was measured experimentally. Then a larger signal of the same duration was paired with the first in an amplitude discrimination experiment, and its detectability relative to the first was determined. Then the detectability scale value assigned to the larger signal is the detectability of the smaller, plus the detectability of the difference between the first and the second. This gave the scale values marked with open circles in Fig. 1. The worth of such a scaling procedure can be argued by reference to two criteria. The first is to show consistency, for instance by constructing the scale using different-sized step intervals, showing that the obtained scale values lie on the same function, and the second is to show the utility of the obtained function for predicting other relationships. The second of these criteria is the major concern of the present paper.

The lines which define the scales of Fig. 1 were obtained by an averaging process. The scale values for each duration for each observer were superimposed, and an average function was drawn in by eye to the combined function. Then this average function was fitted separately to the scale values for the separate durations. As can be seen, the data are not sufficient to show whether the assumption of parallel functions is correct. For the present the assumption was adequate. The data points represent only 200 to 400 observations, which is not sufficient for finer discrimination of the shape of the curves. The horizontal lines on the curves of Fig. 1 will be discussed below.

LOUDNESS MATCHING

In this experiment the same four observers were run individually in the same sound-treated room. The observer was presented alternately with two tones, about 1 sec apart. Coincident with each of the tones was the flash of a neon light, a separate light marking each of the alternating tones. The duration of the light flashes was

[5] S. M. Pfafflin and M. V. Mathews, "Energy-Detection Model for Monaural Auditory Detection," J. Acoust. Soc. Am. 34, 1842–1853 (1962).

precisely that of the auditory signals. The first signal was fixed at 0.32 sec in duration, and the second varied in duration from one session to the next. An unmarked continuous attenuator was provided, and the observer was instructed to set the attenuator so that the second tone matched the 320-msec tone in loudness. When he was satisfied with the obtained match, he called through an intercom to the control room outside, the signal was disconnected from his earphone and switched to a dummy load, and the signal voltage he had set with his attenuator was read from a Balentine voltmeter. A second attenuator was in series with the one under the observer's control, and this was changed haphazardly from one trial to the next. The observer made five successive matches before the experimental conditions were changed.

Experimental conditions. The observers worked in the same continuous background noise as in the detection experiments. They listened monaurally, using the same earphones as in the detection experiments. The same signal durations were used, ranging from 5.0 to 320 msec. Each of these was paired with a standard duration of 320 msec, which was presented at four different signal-voltage levels. The voltage levels differed between the observers, and an attempt was made to cover the range of the detectability scale for each observer.

The procedure differed from the one traditionally used for such experiments (e.g., Garner[6]). Here it was the *shorter* tone that was under the control of the observer, instead of the longer standard. The more traditional procedure comes from the definition of the loudness of a sound as that value of a 1-kc/sec tone (strictly, a continuous tone) which leads to a judgment of equality. The design of the present experiment dictated the reverse procedure. Logically the two are equivalent, in that the relative subjective intensity of the resultant match can be gotten either way. It is an empirical question whether the same results will be found using the same observers either way. This question was not tested here, although it is a relevant one, and Small *et al.*[7] have presented data which support the assumption of equivalence.

Results. The intensity to which the shorter of the two signals was set in order to match the longer is shown in Fig. 2, in decibels relative to the voltage of the fixed longer tone. Durations are plotted on a logarithmic scale. For each observer four curves are shown, one for each selected voltage of the standard signal. In order to reflect relative loudness, the curves would have to be inverted, but they would have the same form. The spacing of the curves is arbitrary; they should be somewhat closer together in order to reflect the relative voltages accurately, and the spacing would not have

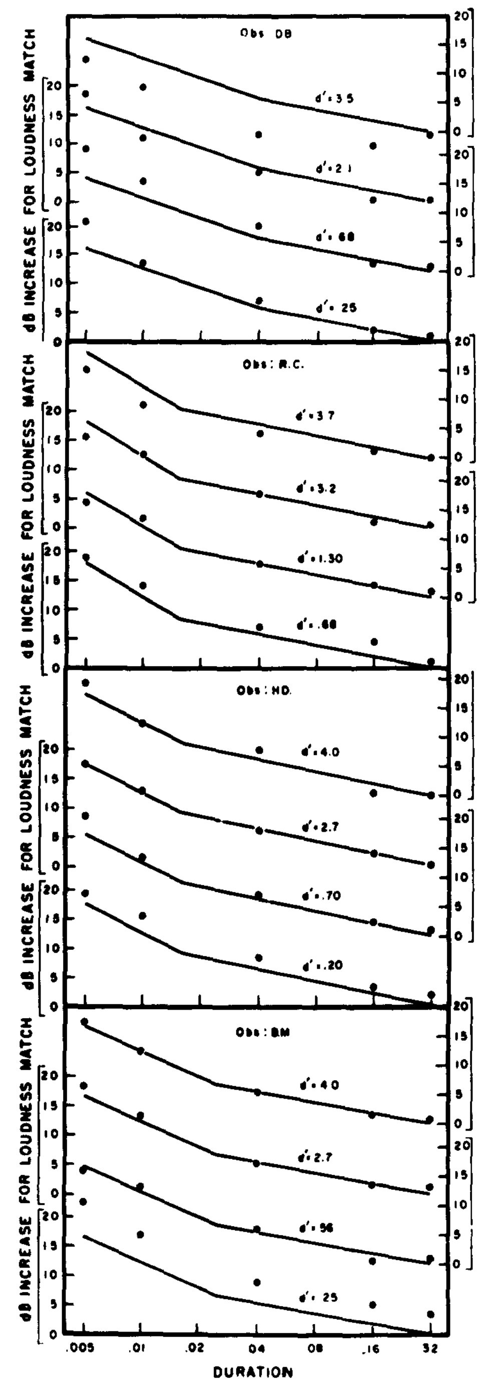

FIG. 2. Results of loudness matching: the relative increase in dB over the 320-msec standard signal for each duration of the variable signal. Duration of the variable is plotted on the abscissa on a logarithmic scale. Each curve has its own ordinate to the side of the graph. The solid lines are predicted matches from the detectability scale. The parameter is the scale value used for prediction.

[6] W. R. Garner, "The Loudness and Loudness Matching of Short Tones," J. Acoust. Soc. Am. 21, 398–403 (1949).

[7] A. M. Small, Jr., J. F. Brandt, and P. G. Cox, "Loudness as a Function of Signal Duration," J. Acoust. Soc. Am. 34, 513–514 (1962).

been even. Each separate curve should be read against its own ordinate value; there are four separate ordinates for each observer.

Prediction of loudness matching. The solid lines on the curves of Fig. 2 represent the theoretical assertion that *the loudness of a signal equals its detectability scale value.* The curves were generated in the following way: a fixed voltage was selected at which to present the longer 0.32-sec signal. The detectability scale value for that signal voltage was read from the left-hand (0.32-sec) curve of Fig. 1, and the prediction of the voltage which would be set to give a loudness match for any other duration was taken to be the signal voltage which gave the same detectability scale value for that duration. The horizontal lines of Fig. 1 determined predictions for the four different levels of the longer standard. The parameter is the scale value corresponding to the lines.

DISCUSSION

As a test of the adequacy of the prediction scheme, an analysis of variance was run using the discrepancy between the mean obtained matches and the predicted values for each observer and each level of the standard. To the degree that significant results come from the statistical test, the prediction is called into question. The only statistically significant variable was the scale-value level. On the average, for the lowest scale values the prediction was 1.2 dB too low, and for the highest scale values the prediction was 2.3 dB too high. The data were stable enough that these small differences were statistically significant.

There was no significant difference between the predicted and obtained matches that could be attributed to the duration of the variable tone. Then the results of the experiment suggest that the form of the duration-loudness function is predicted by the detectability of the signals, but the precise level of the function may not be. This is not unreasonable: at the lowest levels, the detectability of the longer standard tone was quite small, below the detection threshold by most generally accepted criteria. (A d' value 1.40 represents about 75% correct responses in the two-alternative experiment.) "Loudness matching" in this case was difficult for the observers, but they were able to perform the task and to give stable results. It seemed, however, that some of them adopted a somewhat higher criterion (of detectability) for adjusting the variable signals than the one provided. Likewise at the higher signal levels the observers seemed sometimes to adopt a somewhat lower criterion for adjusting the short tone than provided by the fixed-intensity longer tone which alternated with it. In spite of this, the form of the obtained loudness-matching function can be accounted for by the detectability scale.

The form of the predicted functions is also interesting in that it tends to corroborate the functional relation found by Green, Birdsall, and Tanner[1] between the detectability and duration of a tone. The predictions drawn from the psychophysical functions of Fig. 2 can be represented by two straight lines of different slope on the log–log plot. No observer in the prior study showed a break in the function at a duration as large as 0.04 sec, as does *DB*, but the discontinuity falls in the expected range for the other three observers.

This double linear loudness-duration function has not been found by previous investigators. In a recent study which used gated noise signals as stimuli in an adjustment task similar to that used here, Small *et al.*[7] found similar results to previous investigations, linear slopes which showed a loudness change of 12.5 dB per decade change in duration. The slopes of the lines of Fig. 2 are about 6.5 dB per decade for the lower portion of the curves, and about 15.0 dB per decade change in duration for the upper portion, associated with the shortest signals. The finding of two functions may have been a function of the noise level present continuously in the present experiments, for background noise was not present in any of the previous studies. It may also be associated with the change from "tone pitch" to "click pitch" as periodic signals are made shorter, and if this is the case, it should not be expected to occur with noise signals.

In the recent literature two different attacks on experimental problems have been based on measured discrimination. The first of these was a study by Heinemann[8] who showed that differences in the brightness of visual stimuli that result from surrounding stimulus areas with an annulus could be accounted for by invoking the change in relative discriminability which resulted from the presence of the annulus. The second is an article by Taylor,[9] which reports an impressive theoretical rationalle to account for many different visual illusions, with an application to figural aftereffects, the displacement of apparent position after prolonged stimulation. Taylor's basic premise is that the perceived distance between any two points in space is a monotonic increasing function of the discriminability of the difference between them. This postulate is extended to account for many phenomena of figural aftereffect research.

The present experiment, and Heinemann's, show empirical equivalence between sets of operationally distinct behavioral measurements. They do not rest on the assumption of any underlying processes; in fact, Heinemann points out that his results make the interpretation of notions such as central "inhibition" quite complex. Neither do these studies have only the primary aim of establishing the generality of any fixed relation between stimulus and "sensation." In the same way, Taylor derives predicted displacements in figural after-

[8] E. G. Heinemann, "The Relation of Apparent Brightness to the Threshold for Differences in Luminance," J. Exptl. Psychol. 61, 389–399 (1961).

[9] M. M. Taylor, "Figural After-Effects: A Psychophysical Theory of the Displacement Effect," Can. J. Psychol. 16, 247–277 (1962).

effect from separate studies on the discrimination of position. It would seem that discrimination can and does have utility in accounting for perceptual functions of many different kinds.

ACKNOWLEDGMENTS

This research was supported in part under Contract Nonr-248(55) between the Office of Naval Research and The Johns Hopkins University. This is Report No. 26 under that contract. Reproduction in whole or in part is permitted for any purpose of the United States Government. The author was much aided in this research by Public Health Service postdoctoral research Fellowship No. 12,140 from the National Institute of Mental Health.